

DEVELOPMENT OF A SHORT QUESTIONNAIRE FOR USE IN EPIDEMIOLOGICAL STUDIES OF DEPRESSION IN CHILDREN AND ADOLESCENTS

ADRIAN ANGOLD* MRCPsych, ELIZABETH J. COSTELLO PhD and STEPHEN C. MESSER PhD
*Developmental Epidemiology Program, Duke University Medical Center, Box 3454, Durham,
North Carolina 27710-3454, USA*

ANDREW PICKLES PhD, FRANCES WINDER BSc and DANA SILVER BSc
*Department of Child & Adolescent Psychiatry, MRC Child Psychiatry Unit, Institute of Psychiatry,
16 DeCrespigny Park, Denmark Hill, London SE5 8AF, UK*

SUMMARY

The purpose of the present study was to describe the development and assess the psychometric properties of the Short Mood and Feelings Questionnaire (SMFQ). The SMFQ is a brief, easy-to-administer, self-report measure of childhood and adolescent depression, designed for the rapid evaluation of core depressive symptomatology or for use in epidemiological studies. The SMFQ's content and criterion-related validity were examined in a sample of 173 8-16 year-olds, comprised of both psychiatric and unselected pediatric controls. Results revealed substantial correlations between the SMFQ, the Children's Depression Inventory (CDI) and the Diagnostic Interview Schedule for Children (DISC) depression scale. The SMFQ successfully discriminated the clinically-referred psychiatric subjects from the pediatric controls. Within the pediatric (general population) sample, the SMFQ discriminated DISC-diagnosed children with depressive disorder from non-depressed subjects. Exploratory factor analyses, along with a high internal consistency, suggested that the SMFQ was a unifactorial scale. In sum, the SMFQ appears to be a promising tool for both the swift assessment of core depressive symptomatology and as a screening measure for depression in child psychiatric epidemiological studies.

KEYWORDS—depression; children and adolescents; assessment

INTRODUCTION

Structured psychiatric interviews are expensive and time-consuming, but they are the central pillar of the clinical and research psychiatric assessment of both adults and children. With children, these problems are exacerbated by the need, in most cases, for an interview with both the child concerned and a parent (Weissman *et al.*, 1987). Childhood depression has a six-month prevalence of only 2-5% (Anderson *et al.*, 1987; Boyd and Weissman 1981; Fleming and Offord, 1990; Guyer *et al.*, 1989; Kashani *et al.*, 1983; McGee *et al.*, 1990; McGee and Williams, 1988), which means that most children interviewed in general population studies of depression will not

have the disorder, so that a brief screening questionnaire that identified children likely to have the disorder at interview would be very useful. Such a questionnaire should be available in both self-report and parent-report forms. It also needs to be as short and simple as possible, both to reduce subject burden and to maximize further participation in later stages of a project. We have pointed out elsewhere that very high degrees of sensitivity and specificity cannot be expected from such instruments and that they might be better regarded as 'nets' than as 'screens' (Costello and Angold, 1988). Even though nets will miss a number of cases of interest and pick up a good deal of other material, they can be expected to increase the 'concentration' of cases in the interviewed sample sufficiently to be cost-effective.

Such a questionnaire could also serve a useful purpose in providing for the rapid assessment of

*Author to whom correspondence should be addressed.

core depressive symptomatology in clinical and epidemiological studies where depression is not the primary focus of attention, but where at least a rough index of depression is required. A further potential use is as a brief measure suitable for more frequent follow-ups than the usual annual or semi-annual major assessments employed in many psychiatric epidemiological studies.

However, a net or screen that does not work would not be cost-effective under any circumstances and may prove counter-productive (Kashani *et al.*, 1983). Attention to some practical psychometric considerations (particularly the important distinction between criterion-related and content validity) suggests that many of the available depression questionnaires for children and adolescents are likely to be suboptimal in their netting properties (Costello and Angold, 1988; Kazdin, 1987; Reynolds, 1982). In this paper, therefore, we discuss some of the important design considerations for a netting questionnaire and describe the development and psychometric properties of the Short Mood and Feelings Questionnaire (SMFQ), which we have developed for use in future epidemiological studies.

CRITERION-RELATED VERSUS CONTENT VALIDITY

Researchers are often interested in at least two aspects of the phenomenology of depression. The first is the range of symptoms displayed by the subjects under study. They therefore require that their measures adequately reflect the range of phenomena that are associated with the disorder in which they are interested.

Depressive symptomatology is quite wide-ranging, and most depression questionnaires include a variety of items covering areas as diverse as appetite disturbance, guilt and suicidal thinking. The degree to which a questionnaire, interview or test covers the full range of the material of interest is often referred to as its 'content validity'.

However, researchers often want to take a step beyond the symptoms and to define disorders or diseases. If a questionnaire is being used to select subjects who are likely to have a particular disorder, as defined by some criterion measure, then the degree to which it does so is referred to as the instrument's 'criterion-related validity'. If this ability to detect disorder is the only consideration,

then the best instrument will usually be the one that performs the task of selecting cases and rejecting non-cases most efficiently and in the shortest time; in other words with as few items as possible. 'Concurrent validity' (a subset of criterion validity) refers to the degree to which a new measure agrees (or correlates) with assessments of the same construct by other (usually pre-existing) measures.

As Cronbach (1970) has pointed out, the requirements of content validity and criterion validity are often contradictory, since a measure that covers the area fully is likely to be relatively long and may be a poor predictor of a particular criterion of interest, while one that is short but highly predictive may provide poor content coverage. The Children's Depression Inventory (CDI) (Kovacs, 1983; Saylor *et al.*, 1984a, 1984b) and the Children's Depression Scale (CDS) (Lang and Tisher, 1978; Rotundo and Hensley, 1984; Kazdin, 1987), which are possible candidates for the role of first-stage netting questionnaire in epidemiological studies fall between the two stools of content- and criterion-related validity, combining moderate coverage of the area with relatively low predictive power for the diagnosis of depression (Costello and Angold, 1988). For instance, they have been reported to have very high levels of internal reliability (Cronbach's alphas frequently > 0.9), suggesting a great deal of redundancy in the items as measures of the latent depression variable (Boyle, 1985). In other words, these scales are probably longer than they need to be for netting purposes. We therefore set out to produce a netting questionnaire (the Short Mood and Feelings Questionnaire, SMFQ) that was empirically designed to minimize length and maximize criterion validity.

METHOD

Initial item selection

We began by selecting 30 symptom items (see Table 2) reflecting current clinical and taxonomic thinking about childhood depression (e.g. Ryan *et al.*, 1987; American Psychiatric Association, 1980), which tapped affective, cognitive, vegetative and suicidal aspects of depression. Symptoms such as enuresis, wandering behavior and school phobia, whereas they might be related to depression in childhood and adolescence and are contained in some other depression scales (Angold, 1988), are

Table 1. Sample characteristics

		Pediatric group n (%)	Psychiatric group n (%)
Sex	Boys	54 (43)	33 (69)
	Girls	71 (57)	15 (31)
Age (years)	6-11	125	23
	12-17	0	25
Race	White	101 (84)	31 (62)
	Black	20 (16)	16 (32)
	Missing	4	1

not central to the concept of depression as it is usually understood and were not included. A parallel version was written for parents, containing the same items, with one addition that made sense only as something reported about the child by another person ('S/he was not as happy as usual even when you praised or rewarded him/her'). In order to make the questionnaire as simple as possible, each item consisted of a single sentence to which the subject could respond 'Not true', 'Sometimes' or 'True'.

We refer to this initial item pool as the Mood and Feelings Questionnaire (MFQ). The suffix C indicates the child self-report version, while P indicates the parent-report version.

Samples

Symptom scores were studied in two samples that were already available to us through their involvement in other studies, one referred for psychiatric services and one not so referred.

Psychiatric group. This group comprised 48 consecutive child psychiatric outpatient referrals attending the Western Psychiatric Institute and Clinic in Pittsburgh and aged between 6 and 17 years.

Pediatric group. This group comprised 125 6-11 year-olds (only one was younger than seven) who had been brought to the primary care pediatric clinic of a large Health Maintenance Organization for a wide variety of general health problems. This group was a consecutive subsample

of children involved in a study of psychiatric morbidity and service use that included around 80% of children enrolled in the pediatric group. Costello *et al.* (1988) have shown that this group was a close approximation to a general population sample both in terms of its demographics and patterns of psychiatric disorder.

Table 1 shows the demographic characteristics of the two groups.

Measures

In addition to the initial item pool, the Children's Depression Inventory (CDI) (Kovacs, 1983) was used as one point of comparison. The CDI is currently the most widely used depression questionnaire for children, designed with a format and content similar to that of the Beck Depression Inventory (BDI) (Beck and Beamesderfer, 1974). Each item consists of three statements about a symptom area, encompassing different levels of severity (for example; I hate myself; I do not like myself; I like myself.) The child endorses the one closest to how he or she has felt or thought during the preceding 2 weeks. Though this approach works well enough for many purposes, it is somewhat more complex than the method of questioning adopted for the SMFQ, in that it requires the subject to compare three statements for each item, rather than making a decision about a single statement.

In the pediatric group, the criterion used in making a diagnosis of DSM-III depression was a detailed psychiatric assessment of the child using the Diagnostic Interview Schedule for Children (DISC) (Costello *et al.*, 1982). This highly structured instrument consists of two parallel interviews about the child's emotional and behavioral problems: the DISC-C for children and the DISC-P for parents and caretakers. Responses to the DISC questions are coded as 0 (no), 1 (sometimes or maybe) or 2 (yes). They are computer scored to yield symptom scores, and also DSM-III diagnoses, at two levels of severity, for a wide range of psychiatric disorders. Level 1 ('possible' diagnosis) is a precise operationalization of the DSM-III criteria. A weighted summary score generated by the DISC diagnostic algorithms was derived from the responses to the DISC and is referred to as the Total DISC Depression Score. This consisted of the sum of the scores on each of the depression subsections of the DISC (affective, cognitive,

Table 2. Child initial item pool: item total predictive power

MFQ items	Total MFQ-C score given item score (n)			PE ¹	SPE ²	p	Correlation ³ item – total
	0	1	2				
Miserable or unhappy	5.6 (55)	13.7 (61)	19.3 (8)	0.13	0.68	0.0001	0.52
Didn't enjoy anything	9.3 (106)	19.0 (18)	31.0 (1)	0.11	0.56	0.0001	0.42
Less hungry	7.6 (75)	13.6 (28)	16.1 (22)	0.09	0.45	0.0001	0.42
Ate more	9.0 (68)	10.9 (37)	14.0 (19)	0.04	0.20	0.03	0.12
Tired	6.2 (67)	14.2 (34)	17.0 (22)	0.12	0.59	0.0001	0.48
Restless	6.2 (71)	12.3 (30)	20.4 (24)	0.17	0.84	0.0001	0.59
No good	8.1 (102)	22.0 (20)	28.1 (3)	0.21	1.04	0.0001	0.52
Blamed self	7.6 (85)	14.2 (15)	17.9 (25)	0.18	0.59	0.0001	0.42
Indecisive	5.5 (54)	11.7 (42)	16.6 (29)	0.12	0.60	0.0001	0.43
Irritable	7.5 (79)	15.2 (33)	16.8 (13)	0.10	0.51	0.0001	0.43
Talking less	7.0 (75)	15.6 (35)	15.8 (15)	0.10	0.52	0.0001	0.43
Cried a lot	9.3 (104)	17.4 (17)	27.0 (3)	0.11	0.54	0.0001	0.46
No good in future	9.1 (106)	18.1 (14)	26.8 (4)	0.13	0.64	0.0001	0.45
Not worth living	8.5 (105)	19.3 (11)	26.8 (9)	0.20	0.98	0.0001	0.56
Thoughts of death	8.9 (107)	20.8 (13)	25.8 (5)	0.16	0.82	0.0001	0.44
Better off without	8.9 (109)	20.6 (8)	26.5 (8)	0.19	0.96	0.0001	0.47
Suicidal thoughts	9.9 (117)	27.7 (6)	21.5 (2)	0.16	0.83	0.0003	0.38
Not see friends	9.7 (105)	19.7 (10)	20.9 (4)	0.11	0.55	0.0003	0.33
Poor concentration	6.7 (81)	15.2 (34)	25.9 (10)	0.19	0.94	0.0001	0.66
Bad things happen	7.2 (89)	16.1 (24)	25.1 (12)	0.20	1.0	0.0001	0.63
Hated myself	8.4 (102)	20.7 (16)	22.7 (7)	0.15	0.77	0.0001	0.47
Bad person	8.1 (101)	21.4 (21)	27.3 (3)	0.19	0.98	0.0001	0.55
Looked ugly	9.1 (95)	13.7 (18)	18.3 (12)	0.07	0.37	0.0009	0.32
Lonely	7.2 (73)	12.3 (39)	23.5 (13)	0.14	0.69	0.0001	0.56
Unloved	8.4 (103)	17.7 (11)	25.9 (10)	0.18	0.91	0.0001	0.55
Never be as good	8.0 (90)	15.3 (28)	26.3 (7)	0.13	0.67	0.0001	0.54
Did everything wrong	7.6 (97)	19.5 (24)	33.3 (4)	0.25	1.25	0.0001	0.60
Poor sleep	7.3 (80)	12.8 (21)	18.8 (24)	0.13	0.63	0.0001	0.49
Slept more	4.5 (119)	10.0 (30)	18.3 (17)3	0.06	0.31	0.0019	0.30

¹PE, parameter estimate; ²SPE, standardized parameter estimate; ³Pearson product-moment correlation coefficient.

vegetative and suicidal) each standardized to range between 0 and 50.

Procedure

The initial version of the MFQ was completed as part of a larger psychiatric assessment of the pediatric and psychiatric groups. The children either completed the form on their own or had it read to them, depending upon their age and reading competence. The answers recorded were the subjects' self reports and not the examiners' opinions about them. One parent was also asked to complete the parent version of the questionnaire

about the target child. The children also completed the CDI. The order of administration of the MFQ and CDI was randomized within each sample. The parents and children in the pediatric group were then seen separately by different research interviewers, blind to the MFQ responses, who completed the DISC with them. The psychiatric sample was also seen by an experienced psychologist or social worker, and by a child psychiatrist. However, no structured diagnostic interview was performed and the well-known unreliability of clinical diagnoses (see, e.g. Cantwell, 1988) precluded comparisons of MFQ scores between different diagnostic groups within this sample.

Table 3. Parent initial item pool: item total predictive power

MFQ items	Total MFQ-P score given item score (n)			PE ¹	SPE ²	p	Correlation ³ item-total
	0	1	2				
Miserable or unhappy	2.0 (83)	6.9 (32)	16.6 (8)	0.35	1.1	0.0001	0.74
Didn't enjoy anything	3.9 (115)	12.0 (10)		0.14	0.45	0.0021	0.53
Less hungry	3.9 (112)	9.8 (13)		0.11	0.36	0.0059	0.38
Ate more	3.6 (101)	7.4 (18)	8.3 (6)	0.10	0.32	0.0026	0.34
Tired	3.7 (104)	7.5 (19)	17.5 (2)	0.12	0.38	0.0010	0.50
Restless	2.4 (96)	8.0 (21)	17.9 (8)	0.37	1.15	0.0001	0.69
No good	3.2 (111)	14.2 (13)	30.0 (1)	0.37	1.18	0.0001	0.71
Blamed self	3.9 (113)	9.4 (11)	30.0 (1)	0.16	0.54	0.0001	0.48
Indecisive	3.0 (81)	6.4 (40)	9.8 (4)	0.10	0.33	0.0015	0.52
Irritable	1.6 (63)	5.6 (53)	9.8 (9)	0.22	0.68	0.0001	0.53
Talking less	3.7 (109)	8.7 (15)	35.0 (1)	0.16	0.54	0.0001	0.60
Cried a lot	2.9 (103)	10.9 (18)	15.5 (4)	0.22	0.72	0.0001	0.54
No good in future	3.9 (118)	13.6 (6)	35.0 (1)	0.25	0.81	0.0001	0.58
Not worth living	4.1 (119)	23.4 (3)		0.22	0.71	0.0030	0.71
Thoughts of death	4.3 (118)	9.8 (6)	8.0 (1)	0.08	0.28	0.0472	0.49
Better off without	3.8 (117)	15.6 (8)		0.20	0.65	0.0007	0.63
Suicidal thoughts	4.5 (123)	13.3 (2)		0.10	0.35	0.0855	0.46
Not see friends	4.3 (118)	10.4 (7)		0.10	0.32	0.0237	0.32
Poor concentration	2.9 (99)	7.6 (21)	20.6 (5)	0.24	0.77	0.0001	0.57
Bad things happen	4.2 (120)	13.8 (4)	19.7 (1)	0.22	0.47	0.0022	0.45
Hated myself	3.6 (116)	13.0 (7)	32.5 (2)	0.35	1.15	0.0001	0.69
Bad person							
Looked ugly	3.2 (106)	10.3 (17)	23.5 (2)	0.22	0.71	0.0001	0.46
Lonely	2.7 (96)	8.5 (27)	32.5 (2)	0.31	0.98	0.0001	0.72
Unloved	3.0 (107)	12.4 (17)	30.0 (1)	0.32	1.04	0.0001	0.70
Never be as good	3.3 (109)	11.6 (15)	35.0 (1)	0.28	0.91	0.0001	0.67
Did everything wrong							
Poor sleep	3.6 (110)	8.1 (11)	18.3 (3)	0.15	0.47	0.0002	0.57
Slept more	4.5 (119)	7.4 (6)		0.05	0.18	0.2631	0.31
Not cheered up	3.5 (115)	15.9 (9)	30.0 (1)	0.34	1.11	0.0001	0.70

¹PE, parameter estimate; ²SPE, standardized parameter estimate; ³Pearson product-moment correlation coefficient.

RESULTS

The analyses broadly fall into three parts: (1) assessments of the internal structure of the initial item pool; (2) an assessment of the criterion-related validity of the initial item pool, as measured by its ability to predict clinical group status, and in the form of comparisons with the subjects' DISC scale score and CDI responses; and (3) an assessment of the performance of the subset of empirically derived screening items emerging from (1) and (2).

Content validity of the MFQ

In one sense, content validity is simply determined by looking at the items in the scale

and noting that they do indeed cover the depressive phenomenology embodied in both the DSM and ICD diagnostic systems. However, we have extended our consideration of this concept somewhat to include our findings on the acceptability of the instruments in use.

It was our impression that, as part of a long assessment, the MFQ functioned well with both the younger children and the oldest subjects in the study. Its format presented no obvious disadvantages and seemed simpler in use than that of the CDI. The parent version was also quite satisfactory in these respects. However, as the result of a typographical error while creating the parent version of the scale, the wording for items 24 ('was a bad person') and 32 ('did everything wrong') in the MFQ-P was incorrect. The results

Table 4. Results from principal components analysis of the parent and child initial item pools

	Item loadings on first component	
	MFQ-C	MFQ-P
Miserable or unhappy	0.49	0.71
Didn't enjoy anything	0.46	0.46
Less hungry	0.40	0.28
Ate more	0.16	0.24
Tired	0.48	0.44
Restless	0.64	0.74
No good	0.72	0.74
Blamed self	0.49	0.44
Indecisive	0.50	0.34
Irritable	0.42	0.56
Talking less	0.39	0.52
Cried a lot	0.46	0.65
No good in future	0.52	0.62
Not worth living	0.71	0.54
Thoughts of death	0.64	0.13
Better off without	0.71	0.58
Suicidal thoughts	0.50	0.17
Not see friends	0.34	0.28
Poor concentration	0.68	0.63
Bad things happen	0.69	0.34
Hated myself	0.62	0.78
Bad person	0.67	—
Looked ugly	0.40	0.64
Lonely	0.61	0.72
Unloved	0.69	0.72
Never be as good	0.67	0.67
Did everything wrong	0.75	—
Poor sleep	0.57	0.47
Slept more	0.30	0.06
Not cheered up		0.74

for these items are therefore excluded from all analyses of the parent data.

Internal consistency of the MFQ item pool

As we expected, initial item pools for both parents and children had high internal reliabilities (Cronback's $\alpha = 0.90$ for both), suggesting a good deal of redundancy in the measurement of the latent depression variable. This finding encouraged our search for the subgroup of items that would be most closely linked to the depression variable measured by the scale, so that we could reduce redundancy and increase the criterion-related performance of the final SMFQ.

Individual items as predictors of total MFQ item pool score

The most useful items in a homogeneous scale are those that are strongly associated with elevations in the total score. We compared the total initial item pool scores of those who reported a particular item present with the total initial item pool scores of those who reported that item absent. The total initial item pool score was adjusted to take account of the contribution of the item under consideration by subtracting each subject's score on that item from his/her total initial item pool score. Tables 2 and 3 show the mean total MFQ score for those whose reports contained a 0, 1 or 2 on each item for parent and child reports respectively (where the n values in columns 0, 1 and 2 do not sum to 125, data for that item were missing for one or more cases). Since the initial item pool scores were far from being normally distributed (all had 'reverse J' shaped distributions), the parameter estimates and significance values shown in Tables 2 and 3 represent the findings from maximum likelihood logistic regressions for ordinal response data of total MFQ score on item score.

The first point to note is that, overall, the rates of positively reported symptoms were much higher for the child self-reports than for the parent reports. In particular, parents rarely used codings of 2. Secondly, nearly all of the items were substantially associated with the adjusted total score in both the parent and child data. Another way of looking at the relationships between individual items and the total scale scores is to compute item-total correlations, with the contribution of the item under consideration removed from the total. Pearson correlation coefficients are also shown in Tables 2 and 3.

Factor analysis

The methods of assessing the relationships between individual items and the scale's overall measure of depression discussed so far have used the total score minus the score on the item under consideration as the overall measure. However, it may be more appropriate to look at individual item loadings on the latent variable (or variables) measured by the scale as a whole using factor analysis, since this will exclude a certain amount of error variance that is included in the total score.

Table 5. Initial item pool: items as predictors of sample membership

MFQ items	Child self-reports		Parent reports	
	OR	<i>p</i>	OR	<i>p</i>
Miserable or unhappy	1.9	0.067	3.5	<0.001
Didn't enjoy anything	2.9	0.029	6.1	<0.001
Less hungry	1.4	0.213	3.7	0.001
Ate more	0.5	0.100	1.5	0.125
Tired	1.3	0.208	2.6	0.002
Restless	1.3	0.185	2.6	<0.001
No good	1.2	0.555	2.8	0.007
Blamed self	1.0	0.817	3.1	0.005
Indecisive	0.8	0.337	2.0	0.013
Irritable	2.4	<0.001	2.1	0.003
Talking less	1.5	0.082	5.6	<0.001
Cried a lot	2.5	0.002	1.5	0.186
No good in future	1.5	0.214	4.7	<0.001
Not worth living	1.2	0.608	12.0	<0.001
Thoughts of death	2.1	0.012	4.9	0.001
Better off without	1.8	0.022	7.1	<0.001
Suicidal thoughts	2.5	0.019	7.5	0.016
Not see friends	2.0	0.029	4.2	0.004
Poor concentration	2.1	0.003	2.8	<0.001
Bad things happen	1.1	0.628	5.5	<0.001
Hated myself	1.1	0.868	2.5	0.024
Bad person	1.8	0.083		
Looked ugly	0.9	0.815	1.9	0.067
Lonely	1.2	0.388	2.9	<0.001
Unloved	1.2	0.488	3.9	<0.001
Never be as good	1.3	0.322	2.3	0.029
Did everything wrong	1.3	0.460		
Poor sleep	1.0	0.975	2.2	0.020
Slept more	0.8	0.296	2.0	0.205
Not cheered up			3.9	<0.001

Given the item distributions, maximum likelihood (ML) factor analysis would have been a method of choice here, but it resulted in an ultra-Heywood solution (communalities summed to more than 1.0—a common problem with ML factor analysis) with the child self-report data. Other forms of factor analysis resulted in very similar solutions, and so we report the results of the familiar principal components analysis here. With the parent-report data, the results of ML factor analysis were very similar to those from principal components analysis, and again the latter is reported here to maximize comparability

Table 6. CDI items as predictors of pediatric versus psychiatric group status

Item	OR	<i>p</i>
1	2.4	0.041
2	1.2	0.574
3	1.7	0.223
4	1.1	0.693
5	2.4	0.126
6	1.6	0.110
7	1.8	0.215
8	1.1	0.835
9	1.8	0.149
10	1.1	0.869
11	1.6	0.098
12	1.5	0.374
13	0.8	0.485
14	1.2	0.526
15	2.0	0.009
16	1.4	0.199
17	1.1	0.774
18	0.9	0.678
19	0.9	0.843
20	1.5	0.186
21	1.8	0.052
22	1.9	0.054
23	1.7	0.065
24	1.2	0.428
25	1.5	0.337
26	3.6	<0.001
27	2.4	0.015

between the results for parents and children (see Table 4).

In both the parent and child data, a large first factor, involving moderate to high loadings on most items, stood out. The second factor in an unrotated solution appeared, in both cases, mostly to reflect the absence of the items with the highest loadings on the first factor, but, like the other small factors, was unstable under a variety of rotation techniques, whereas the first factor remained heavily weighted towards the affective and cognitive items.

Criterion-related validity

Prediction of clinical status and depression status by the MFQ items. If items are to be used for prediction, it is obviously necessary to demonstrate that they do, indeed, have predictive power for the criterion variables of interest. Two

Table 7. Initial item pool: items as predictors of DISC depression diagnosis

	Self-reports		Parent reports	
	OR	p	OR	p
Miserable or unhappy	5.7	0.004	3.2	0.011
Didn't enjoy anything	2.2	0.248	6.6	0.017
Less hungry	1.2	0.598	1.0	0.966
Ate more	1.7	0.239	1.7	0.327
Tired	4.7	0.001	2.5	0.113
Restless	2.5	0.024	2.1	0.098
No good	4.9	0.003	4.5	0.020
Blamed self	1.5	0.261	3.4	0.075
Indecisive	1.7	0.219	1.5	0.488
Irritable	2.3	0.047	1.4	0.488
Talking less	2.1	0.084	9.1	<0.001
Cried a lot	1.1	0.915	3.3	0.015
No good in future	2.5	0.090	2.5	0.090
Not worth living	2.8	0.014	6.1	0.155
Thoughts of death	4.1	0.003	—*	—*
Better off without	3.1	0.006	9.4	0.007
Suicidal thoughts	2.2	0.252	—*	—*
Not see friends	3.2	0.021	2.1	0.536
Poor concentration	2.7	0.023	2.2	0.123
Bad things happen	3.2	0.005	—*	—*
Hated myself	3.0	0.012	6.9	0.003
Bad person	3.6	0.014	—	—
Looked ugly	1.7	0.189	3.8	0.018
Lonely	1.5	0.385	2.6	0.089
Unloved	4.1	<0.001	5.2	0.009
Never be as good	3.1	0.015	9.1	<0.001
Did everything wrong	3.7	0.009	—	—
Poor sleep	3.0	0.005	2.9	0.053
Slept more	1.2	0.689	—*	—*
Not cheered up			4.0	0.048

*Solution would not converge.

criteria were available to us for this purpose: clinical status (pediatric versus psychiatric) and depression status as measured by the DISC in the pediatric group.

Comparisons of individual item frequencies by study group. Table 5 shows the odds ratios (OR) and significance levels from maximum likelihood logistic regressions of individual item scores on group status (pediatric versus psychiatric). An

odds ratio of 2 indicates that for each increment of 1 (from a score of 0 to 1 or from a score of 1 to 2) the odds of being in the psychiatric group doubled. The parent items proved to be far better predictors of group status than the child self-report items, which is perhaps not surprising, given that parents are usually responsible for bringing their children for treatment. However, several of the child self-report items were associated with substantial increases in the odds of being in the psychiatric group.

In order to determine whether the poor item-by-item prediction of clinical status was a particular feature of the MFQ-C or a more general feature of self-reports of depressive symptoms, we conducted a similar analysis on the items of the CDI (Table 6). The pattern of results proved to be very similar for the CDI and initial item pool.

Prediction of DISC depression status in the pediatric group. The following diagnoses generated by the DISC were pooled as 'depression diagnoses': dysthymia; major depression; manic-depressive disorder; cyclothymia. There were six parent reports of possible or probable depression diagnoses from the 125 subjects in the pediatric sample and five possible or probable diagnoses from the child reports, in 10 children altogether.

To determine the ability of individual items to predict DISC depression status we conducted a further series of logistic regressions of item scores on DISC depression status. The results of these analyses are shown in Table 7. In this case, the child self-reports proved to be better predictors of depression status than the parent reports, even though both parent and child DISC results were included in the determination of depression status.

Selection of items for the SMFQ

The foregoing analyses were used as the basis for selecting a subgroup of items to serve as the SMFQ. The items selected were the 13 items (see Table 8), which had performed well in a variety of analyses. This group was heavily weighted towards the affective and cognitive items in the MFQ item pool, but it also included tiredness, restlessness and poor concentration.

The remainder of this paper deals with the psychometric properties of the SMFQ. We wanted

Table 8. First factor loadings of SMFQ-C and SMFQ-P items

	SMFQ-C	SMFQ-P
Miserable or unhappy	0.45	0.71
Didn't enjoy anything	0.36	0.41
Tired	0.46	0.31
Restless	0.54	0.73
No good	0.68	0.80
Cried a lot	0.46	0.61
Poor concentration	0.58	0.57
Hated myself	0.62	0.76
Bad person	0.72	—
Lonely	0.48	0.62
Unloved	0.68	0.82
Never be as good	0.65	0.56
Did everything wrong	0.78	—

to have identical items in the parent and child versions to allow direct comparisons between parent and child reports of individual items in the future, but since two of the items which performed well in the child reports were those with faulty questions in the parent version, the analyses of the parent data include only the 11 usable parent items.

Internal structure of the SMFQ

Internal reliability. Coefficient α was 0.85 for the SMFQ-C and 0.87 for the SMFQ-P. These levels suggest that both scales could be shortened still further. However, at this stage, given the limited size of our samples, we wished to avoid narrowing the scale too severely, since this would carry a risk of reducing the replicability of our results in other populations.

Maximum likelihood factor analysis. Given that one contributor to our decisions about items for inclusion in the SMFQ was a high item loading on the first factor from the principal components analysis of the original MFQ item pool, we expected that the SMFQ would be a unifactorial scale with high item loadings on the first principal factor. This is exactly the pattern that emerged (see Table 8). For both the SMFQ-C and SMFQ-P only a single factor had an eigenvalue greater than 1 (7.98 and 9.92, respectively).

Table 9. Intercorrelations among depression scale scores

	SMFQ-C	SMFQ-P	MFQ-C	MFQ-P	CDI	DISC-C
SMFQ-C						
SMFQ-P	0.30					
MFQ-C	0.96	0.34				
MFQ-P	0.29	0.91	0.25			
CDI	0.67	0.31	0.62	0.24		
DISC-C	0.65	0.28	0.58	0.19	0.48	
DISC-P	0.32	0.43	0.28	0.40	0.23	0.27

Criterion validity

Comparisons with the CDI and DISC depression scores. Table 9 shows the correlations amongst the child and parent total initial item pool scores (MFQ-EC), the SMFQ-C, the SMFQ-P, the CDI and the DISC depression scores. The correlations between the MFQ and the SMFQ are so high as to indicate that little information has been lost in cutting out over half of the initial item pool items. The SMFQ-C correlates moderately highly with both the CDI score and the DISC depression score, and its correlation with the latter is similar to that of the CDI. Correlations between parent and child reports are at the expected level of around 0.2–0.4 for all the scales.

Clinical status. Table 10 shows the mean scores of the pediatric and psychiatric groups on the various depression scales. The SMFQ-P and SMFQ-C discriminated clearly between the psychiatrically referred and the non-referred children.

DISC depression diagnoses. It can be seen from Table 11 that all of the scales discriminated between those with a DISC diagnosis of depression and those without. The SMFQ scales did as well as the CDI. The DISC depression scales are included for comparison purposes, and it can be seen that the other scales performed nearly as well as the DISC data on which the diagnoses themselves were based.

Since it was possible that the MFQ might be detecting psychiatric disturbance in general, rather than specifically depression, we then compared

Table 10. Discrimination of clinical group

	OR	<i>p</i>	Pediatric mean (SD)	Psychiatric mean (SD)
SMFQ-C	1.103	0.004	4.68 (4.66)	7.14 (5.19)
SMFQ-P	1.245	<0.001	2.15 (3.27)	5.79 (4.80)
CDI	1.055	0.026	5.58 (6.57)	8.14 (6.97)

Table 11. Discrimination of DISC depression status in pediatric group

	Not depressed on DISC M(SD)	Depressed on DISC M(SD)	OR	<i>p</i>
SMFQ-C	4.17 (4.24)	10.50 (5.56)	1.26	<0.001
SMFQ-P	1.83 (2.56)	5.80 (6.96)	1.26	0.003
CDI	4.86 (5.93)	13.80 (8.23)	1.16	<0.001
DISC-C	6.45 (4.76)	14.72 (7.92)	1.25	<0.001
DISC-P	5.93 (3.66)	15.50 (10.42)	1.30	<0.001

those in the depressed group with the 29 other individuals who had any other diagnosis but not depression. Although the sample sizes are so small, the MFQ-C still significantly discriminated ($OR=1.15$, $p=0.049$), but the MFQ-P did not ($OR=1.09$, $p=0.30$). Thus we have some evidence for the child self-reports that the MFQ is sensitive to depressive symptomatology specifically, rather than just disturbance in general.

How useful are the SMFQ scale scores as predictors of depression?

Receiver operating characteristic (ROC) curves provide a means of examining the efficiency of a screening instrument at different cut-off levels. Fig. 1 shows ROC curves for the parent and child initial item pools, SMFQ-C and SMFQ-P, with DISC depression diagnosis as the criterion. Combined scores on the SMFQ-C and SMFQ-P (i.e. the SMFQ-C score plus the SMFQ-P score), DISC depression score and CDI score are also plotted against DISC diagnosis for comparison. Overall, the child self-reports discriminated depression status better than parent reports, but the combination of parent and child reports from the SMFQ did better than either alone. It can be seen that the combined SMFQ score provided a substantial improvement over random sample

selection. A sensitivity of 70% and specificity of 85% resulted from a cut-off score of 12 or more on the combined scale. However, the 13 items of the SMFQ-C alone achieved 60% sensitivity and 85% specificity at a cut-off score of 8 or more. This translates into a positive predictive power of 80% and a negative predictive value of 68%.

DISCUSSION

The aim of this study was to develop a short screening questionnaire for depression in the general population of children and adolescents. Our results suggest that the SMFQ is a unifactorial scale that taps an underlying construct of general depression similar to those measured by the CDI and the DISC depression scale scores. It appears to be usable with children and adolescents from the age of six to 17, and, at least in 6–11 year-olds, has useful screening properties. It can be administered in five minutes or less and can easily be scored on the spot by an interviewer. Items addressing the affective and cognitive components of depression tended to be the best predictors of depressive status, and are, therefore, heavily represented in the SMFQ.

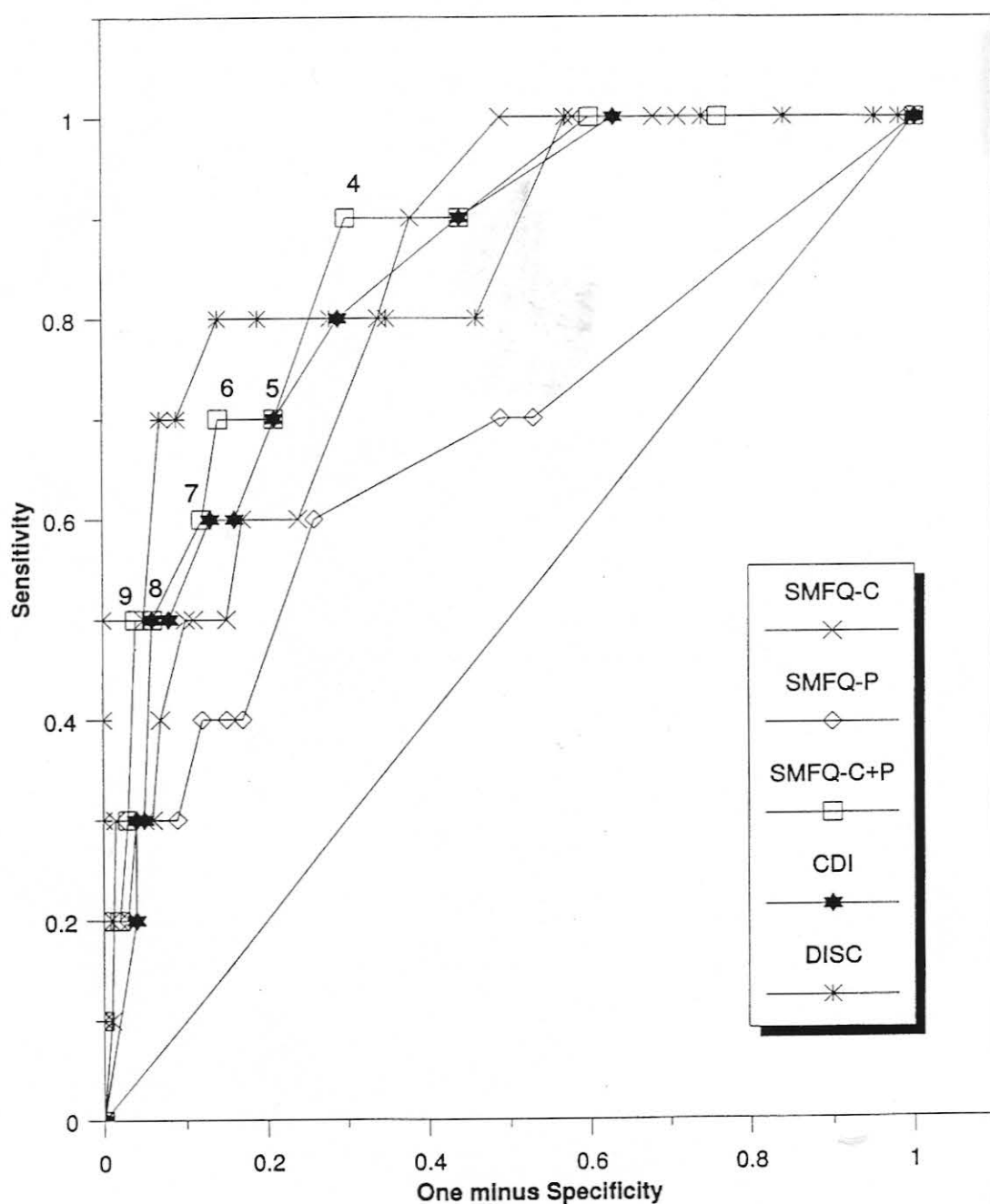


Fig. 1. ROC curves for SMFQ-C, SMFQ-P, SMFQ-C+P, CDI and DISC total depression score against DISC depression diagnosis.

However, it should be noted that these results may overestimate the screening efficiency of the SMFQ. One of the criteria for the inclusion of items in the final scale was success in predicting DISC depression status *in this sample*, and so further testing in other samples will be required to determine how much of the success of the scale is a result of capitalization on chance covariation. On the other hand, given the strength of the correlations between the SMFQ, CDI and DISC depression scale scores, it is unlikely that this represents the whole story. It is striking that, in predicting depression status, both the SMFQ and the CDI were almost as efficient as the DISC depression scores themselves, despite the fact that the DISC depression diagnoses were obviously based on the same data as the depression scores. A further caveat concerns the small numbers involved in the study. In particular, there were few depressed individuals, necessitating the inclusion of DISC 'possible' cases of depression. It would have been preferable to have been able to adapt a more restrictive definition of depression (using, for instance, only DISC 'probable diagnoses'). On the other hand, the finding of significant predictive effects and significant discrimination between those with depression diagnoses and those with other diagnoses with such small numbers is indicative of a large predictive effect size at least for child self-reports.

These results also suggest that a still shorter scale might well be used with little loss of screening efficiency, but further testing of this idea awaits the availability of other data sets. In particular, comparisons of these results with those of studies using different criterion diagnostic instruments are needed.

This study did not include a test-retest stability component, but data from another study (Costello *et al.*, 1991) found 1-week stability in an in-patient adolescent psychiatric population to be quite high (intraclass correlation = 0.75) considering that the time frame of the questionnaire is only 2 weeks. However, as yet, no SMFQ test-retest stability data from a general population sample are available.

Overall, we conclude that the SMFQ shows promise as a short, easily administered screening questionnaire for general population studies of depression.

ACKNOWLEDGEMENT

This work was supported by Contract 278-83-0006(DB) from the National Institute of Mental

Health, by the Medical Research Council and by the Leon Lowenstein Foundation. The authors wish to thank the patients and staff of HealthAmerica Inc., the Western Psychiatric Institute and Clinic, and, in particular, A. J. Costello MD, for help with this study.

REFERENCES

- American Psychiatric Association (1980) *Diagnostic and Statistical Manual of Mental Disorders*, 3rd ed. American Psychiatric Association, Washington, DC.
- Anderson, J. C., Williams, S., McGee, R. and Silva, P. A. (1987) DSM-III disorders in preadolescent children: prevalence in a large sample from the general population *Arch. Gen. Psychiat.* **44**, 69-77.
- Angold, A. (1988) Childhood and adolescent depression I: epidemiological and aetiological aspects. *Br. J. Psychiat.* **152**, 601-617.
- Beck, A. T. and Beamesderfer, A. (1974) Assessment of depression: the depression inventory. In *Psychological Measurements in Psychopharmacology. Modern Problems in Pharmacopsychiatry*, 7th ed. (P. Pichot, ed.), pp. 151-169. Karger and Basel, Paris.
- Boyd, J. H. and Weissman, M. M. (1981) Epidemiology of affective disorders. a re-examination and future directions. *Arch. Gen. Psychiat.* **38**, 1039-1046.
- Boyle, G. J. (1985) Self-report measures of depression: some psychometric considerations. *Br. J. Clin. Psychol.* **24**, 45-59.
- Cantwell, D. P. (1988) DSM-III studies. In *Diagnosis in Child Psychopathology* (M. Rutter, A. Hussain-Tuma and J. S. Lann, eds), pp. 46-52. The Guilford Press, New York.
- Costello, E. J. and Angold, A. (1988) Scales to assess child and adolescent depression: checklist, screens and nets. *J. Am. Acad. Child Adolescent Psychiat.* **27**, 726-737.
- Costello, A. J., Edelbrock, C., Kalas, R., Kessler, M. D. and Klaric, S. H. (1982) *The National Institute of Mental Health Diagnostic Interview Schedule for Children (DISC)*. National Institute of Mental Health: Rockville, MD.
- Costello, E. J., Costello, A. J., Edelbrock, C., Burns, B. J., Dulcan, M. K., Brent, D. and Janiszewski, S. V. (1988) Psychiatric disorders in pediatric and primary care: prevalence and risk factors *Arch. Gen. Psychiat.* **45**, 1107-1116.
- Costello, E. J., Benjamin, R., Angold, A. and Silver, D. (1991) Mood variability in adolescents: a study of depressed, nondepressed and comorbid patients. *J. Affective Disord.* **23**, 199-212.
- Cronbach, L. J. (1970) The two disciplines of scientific psychology. *Am. Psychol.* 671-684.

- Fleming, J. E. and Offord, D. R. (1990) Epidemiology of childhood depressive disorders: a critical review. *J. Am. Acad. Child Adolescent Psychiat.* **29**, 571-580.
- Guyer, B., Lescohier, I., Gallagher, S. S., Hausman, A. and Azzara, C. V. (1989) Intentional injuries among children and adolescents in Massachusetts. *N. Eng. J. Med.* **7**, 1584-1589.
- Kashani, J. H., McGee, R. O., Clarkson, S. E., Anderson, J. C., Walton, L. A., Williams, S., Silva, P. A., Robins, A. J., Cytryn, L. and McKnew, D. H. (1983) Depression in a sample of 9-year-old children: prevalence and associated characteristics. *Arch. Gen. Psychiat.* **40**, 1217-1223.
- Kazdin, A. E. (1987) Children's Depression Scale: validation with child psychiatric inpatients. *J. Child Psychol. Psychiat.* **28**, 29-41.
- Kovacs, M. (1983) *The Children's Depression Inventory: a Self-Rated Depression Scale for School-Aged Youngsters*. University of Pittsburgh School of Medicine, (Unpublished manuscript).
- Lang, M. and Tisher, M. (1978) *Children's Depression Scale*. Australian Council for Educational Research, Victoria, Australia.
- McGee, R. and Williams, S. (1988) A longitudinal study of depression in nine-year-old children. *J. Am. Acad. Child Adolescent Psychiat.* **27**, 342-348.
- McGee, R., Feehan, M., Williams, S., Partridge, F., Silva, P. A. and Kelly, J. (1990) DSM-III disorders in a large sample of adolescents. *J. Am. Acad. Child Adolescent Psychiat.* **27**, 342-348.
- Reynolds, C. R. (1982) Convergent and divergent validity of the Revised Children's Manifest Anxiety Scale. *Educ. Psychol. Measurement* **42**, 1205-1212.
- Rotundo, N. and Hensley, V. R. (1984) The Children's Depression Scale: a study of its validity. *J. Child Psychol. Psychiat.* **26**, 917-927.
- Ryan, N. D., Puig-Antich, J., Cooper, T. B., Rabinovich, H., Ambrosini, P. J., Fried, J., Davies, M., Torres, D. and Suckow, R. F. (1987) Relative safety of single versus divided dose Imipramine in adolescent major depression. *J. Am. Acad. Child Adolescent Psychiat.* **26**, (3), 400-406.
- Saylor, C. F., Finch Jr., A. J., Baskin, C. H., Saylor, C. B., Darnell, G. and Furey, W. (1984a) Children's Depression Inventory: investigation of procedures and correlates. *J. Am. Acad. Child Adolescent Psychiat.* **23**, 626-638.
- Saylor, C. F., Finch, A. J., Spirito, A. and Bennett, B. (1984b) The Children's Depression Inventory: a systematic evaluation of psychometric properties. *J. Consult. Clin. Psychol.* **52**, 955-967.
- Weissman, M. M., Wickramaratne, P., Warner, V., John, K., Prusoff, B. A., Merikangas, K. R., Gammon, G. D., Weisz, J. R. and Cicchetti, D. (1987) Assessing psychiatric disorders in children: discrepancies between mothers' and children's reports. *Arch. Gen. Psychiat.* **44**, 747-753.